Journal of Theoretical and Applied Information Technology
30th November 2014. Vol. 69 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# ARABIC WRITER IDENTIFICATION: A REVIEW OF LITERATURE

**AHMED ABDULLAH AHMED [1], GHAZALI SULONG [2],**
[1], [2] MaGIC-X (Media and Games Innovation Centre of Excellent), Faculty of Computing, Universiti Teknologi Malaysia, Skudai, 81310, Johor Bahru, Malaysia.

E-mail: [1] ah.abdullah85@gmail.com, [2] ghazali@utmspace.edu.my

**ABSTRACT**

In this paper we present a survey of the literature on Arabic writer identification scheme and up-to date techniques employed in identification. The paper begins with an overview of the various writer identification schemes in Arabic and Persian languages. After that, an attempt is made to describe the complex character of Arabic strokes. Previous studies have used a number of Arabic datasets comprised of handwritten text samples that have been used over the years aimed to support handwriting recognition. In automated writer identification, two methods of identification are available text independent and text dependent methods. These are the key focus of this paper. For evaluation purposes we have analyzed the Arabic writer identification scheme by comparing the recognition rate of different methods.

Keywords: *Text independent, text dependent, writer identification, feature extraction, dataset*

## 1. INTRODUCTION

The development in the fields of artificial intelligence and pattern recognition can be applied to the challenging problem of handwriting identification. The Identification of a writers' handwriting is a necessity in the world we live in today due to advancements in technology and its applicability in many fields of human endeavour. The study of writer identification is applicable in many areas such as digital rights management, the financial sphere, solving expert problems in criminology and for decision-making systems to narrowed-down list of identified writers. By combining it with writer verification as an authentication system this can be employed to monitor and regulate access to certain confidential sites or databases with voluminous amounts of documents, notes, forms and meeting minutes which are regularly processed and organized and the identity of the writer would provide additional information.

Automatic identification of the writer from digitized samples of handwriting has been widely studied in the last decade. A variety of writer recognition techniques have been developed for different handwritten scripts (languages). Each script has its own specific characteristics that can be explored to develop a unique approach that works on a given script. This allows a set of writer-specific characteristics to be extracted from written samples from an individual which are compared across different samples with to identify the writer of a given sample.

Handwriting recognition has been identified as having a very close relationship with writer identification [1, 2]. Handwriting recognition involves the elimination of the writer dependent variations in writing and extracts a set of features which can be found in all the different allographs of the same character and permitting its recognition independent of the writing style. Writer identification, on the other hand, depends on the difference of these individual handwriting styles and takes advantage of the variations in characters and words in order to identify the writer of the given sample.

Handwriting identification is the identification of author of a hand written sample of. Handwriting verification is the task to determine whether or not the handwriting is of a given person. Identification and verification of handwriting are visualized in Figure 1 and their comparisons are summarized in Table 1.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

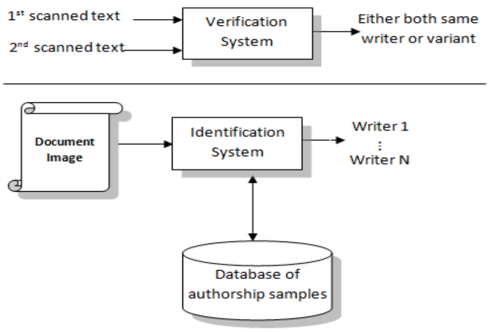


*Figure 1: Writer Verification And Identification Framework [3].*

In handwriting identification, for given handwriting sample x with unknown writer is identified form the samples of handwriting of n known writers. In handwriting verification, for given two handwriting unknown samples x1 and x2 and known hand writing samples of n authors , it is determine that whether x1 and x2 are written by the same person or by two different persons among the n writers. Both models involve classification, with the identification model leading to an n-class problem whereas the verification model leads to a 2-class problem. Both models involve the technique of measuring similarity, or distance, or closeness between two samples. The identification model has the benefit of being able to recognize the writer directly. However, it depends on database which includes information of all the writers in advance. Handwriting verification determines whether two samples of handwriting and are written by the same writer or not.

*Table 1: Comparison Of Identification And Verification Of Handwriting*

| | **Handwriting Identification** | **Handwriting Verification** |
|---|---|---|
| **Similarities** | Measuring similarity between two samples | |
| **Differences** | Giving one sample of handwriting | Giving two samples of handwriting; one the suspicious and other the prototype |
| | n-class classification | 2-class classification, yes or no |
| | decide the writer of the sample | decide whether the two samples are written by same writer or not |

The rest of the paper is organized as follows. Section 2 presents characteristics of the Arabic language. Section 3 demonstrates writer identification framework. Section 4 illustrates the major categories of writer identification domains. Section 5 of the paper discusses the recent research work on offline writer identification as well as the present performance evaluation of various writer identification schemes across Arabic handwritten scripts. Section 6 of the paper presents a brief discussion and conclusion.

## 2. CHARACTERISTICS OF ARABIC SCRIPT:

Arabic writing goes from right to left hand with a total of 28 basic characters in its alphabet construct. It is also comprised of sixteen unique Arabic letters which are composed of one to three dots. The number and the positions of these dots make the difference between these characters otherwise these characters are similar (like ج ,خ and ح). In addition some characters (like ى ,ؤ ,أ ,ك) have zigzag like strokes (Hamza ء). These dots and Hamza are often referred as secondaries and they can be found above the primary part of the character as in the case in ALEF (أ), or below as in the case of BAA(ب), or in the center as in the case of JEEM (ج).

In cursive Arabic scripts, more than two letters may overlap one another without necessarily touching one another. In segmentation–based recognition, this overlapping property could pose some challenges in segmenting words into characters (Figure 2).

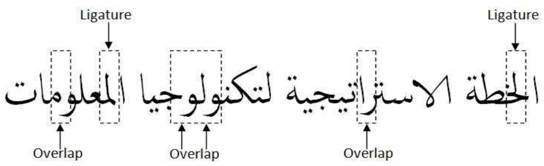


*Figure 2: A sample of the "ligatures" and "overlaps"*

Some Arabic characters use diacritics (a diacritic may be placed above or below the body of the character and this can give the word a different meaning), and many of them do not connect as shown in Figure 3.

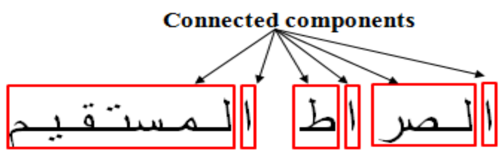


*Figure 3: Characters Depending On The Diacritical Body*

Journal of Theoretical and Applied Information Technology
30th November 2014. Vol. 69 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The position of an Arabic character often effects the shape of the character in the word; a character may be made up of four unique forms depending on its isolation, linkage from the right (ending form), linkage from the left (beginning form), or its connection from both sides (middle form) as presented in Table 2. Word characters often overlap in the vertical direction (not necessarily touch each other). The character size of Arabic words is often fixed both in height and width.

*Table 2: Illustration Of Different Characteristics Of Arabic Script*

| Character Name | Isolated | Initial | Middle | Final | Character Name | Isolated | Initial | Middle | Final |
|---|---|---|---|---|---|---|---|---|---|
| Alif | ا | ا | ـا | ـا | Dhad | ض | ضـ | ـضـ | ـض |
| Ba' | ب | بـ | ـبـ | ـب | Tta' | ط | طـ | ـطـ | ـط |
| Ta' | ت | تـ | ـتـ | ـت | Dha' | ظ | ظـ | ـظـ | ـظ |
| Tha' | ث | ثـ | ـثـ | ـث | A'in | ع | عـ | ـعـ | ـع |
| Jeem | ج | جـ | ـجـ | ـج | Ghain | غ | غـ | ـغـ | ـغ |
| H'a' | ح | حـ | ـحـ | ـح | Fa' | ف | فـ | ـفـ | ـف |
| Kha' | خ | خـ | ـخـ | ـخ | Qaf | ق | قـ | ـقـ | ـق |
| Dal | د | د | ـد | ـد | Kaf | ك | كـ | ـكـ | ـك |
| Thal | ذ | ذ | ـذ | ـذ | Lam | ل | لـ | ـلـ | ـل |
| Rai | ر | ر | ـر | ـر | Meem | م | مـ | ـمـ | ـم |
| Zai | ز | ز | ـز | ـز | Noon | ن | نـ | ـنـ | ـن |
| Seen | س | سـ | ـسـ | ـس | Ha' | ه | هـ | ـهـ | ـه |
| Sheen | ش | شـ | ـشـ | ـش | Waw | و | و | ـو | ـو |
| Sad | ص | صـ | ـصـ | ـص | Ya' | ي | يـ | ـيـ | ـي |

## 3. Writer Identification Framework:

The design of pattern recognition system for off-line text-independent writer identification involves: data acquisition, pre-processing, feature extraction, data representation, and decision making or classification. And these are the main phases in the all framework for writer identification techniques as depicted in Figure 4.

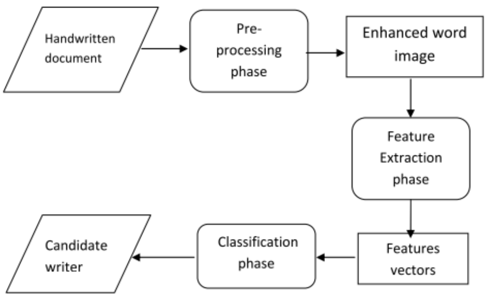


*Figure 4: Writer Identification Framework*

To study the handwriting of different individuals, first, scan the samples into a computer and then automatically obtain handwriting attributes represented as features. In analysis the handwriting samples are compared with collected data and the features extracted from the samples establish the discriminative power of handwriting. Handwriting samples are used to learn the classification task. This classification task has to be learnt from a set of handwriting samples provided by each of the candidates. Once the classification is achieved, a set of samples is used to test the model for its accuracy. For a given test sample of unknown writer, it is determined that the given sample is written by any of the n writers and, if so, to identify the writer. The writer identification procedure utilizes extracted features from the given image and from the labeled prototype images to identify the writer of the test image. In the feature-extraction stage, features are extracted from handwriting and are stored in feature vectors. In the classification stage, the feature vectors are mapped on classes represent the writers.

## 4. MAJOR CATEGORIES OF WRITER IDENTIFICATION DOMAIN:

The terms used in solving the problem of Writer Identification Domain (WID) varies with authorship. Four major categories of WID problems can be identified such as feature extraction, platform, text category, and data specimens. However, most of them share the same goal which is the acquisition of individual features from handwriting that can be used to identify the handwritten authorship. Figure 5 illustrates the classification issues in Writer Identification domain (WID).

Journal of Theoretical and Applied Information Technology
30th November 2014. Vol. 69 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

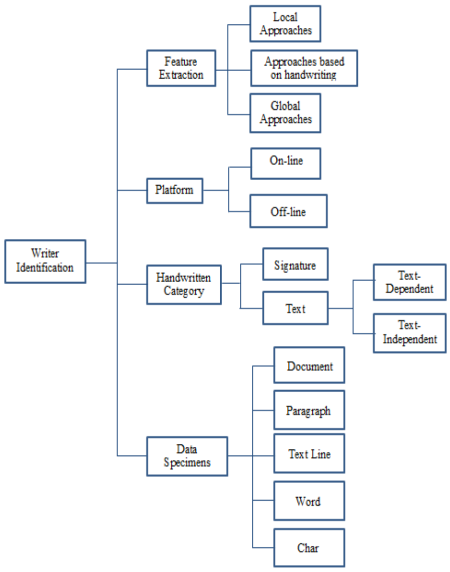


*Figure 5: Classification Issues In Writer Identification.*

## i. FEATURE EXTRACTIONS:

Feature extraction involves the conversion of an input object into vectors comprises of numerical feature sets that enables them to be presented as an object. The information extracted from an object or image in a feature extraction task is called a feature. Local features are the integral parts of objects couple among themselves; while on the other hand, global features are the entire properties that can be ascribed to the object [4]. In a good technique, extracting and selecting of features is very important and has been the focus of research for a long time [5 - 9]. Good features are those that satisfy the two requirements of having small intra-class invariance and large inter-class invariance [9]. Writer identification used for extracting features can be grouped into two major categories.

**A)** Global Features obtained from the text involves manipulation of the text images instead of handwriting. These include Gabor filters and co-occurrence matrices.

**B)** Local Features: Involves manipulation of the handwriting in the form of text images so that several statistical properties can be obtained such as averages of height, width, and legibility of the characters.

## ii. PLATFORM

This category focuses on off-line and on-line systems. The simplest way to distinguish between on-line and off-line writer identification is the input method [10]. Off-line system deals with handwriting images scanned into computer, while on-line system deals with data writing captured by the transducer devices. On-line system has been said to be easy and perform better in accuracy as it relies on the on-line data that is useful as obtained on time order . The dynamic writing process cannot be captured in off line system [11]. Other opinion such as the one by [12] has reported that the performance of off-line system is probably better for WI than signature verification system, since it is generally accepted that the use of several words provides better information as compared to a single signature.

## iii. HANDWRITTEN TEXT CATEGORY

This issue focuses on text-dependent and text-independent identification. Text-dependent identification requires the writer to write the same text which must be fixed accordingly the question document. On the other hand, any text can be used to establish the identification of a writer in text-independent identification. Signature verification or identification falls under text-dependent system because the same signatures are used in the process of training and identification. In the case of handwritten words as used in [12], where one word has been written out for 45 times by the same writer, it has been found to be inapplicable in practical applications such as in determining a crime suspect or in any identification procedure of forensic science [13]. Thus, researchers prefer the text-independent identification such as the one's featured in the works of Dynamic Features [14]; Expert System [15]; Personal Identification Texture Analysis [16]; HMM Based Recognizer [17]. In Table 3 the difference between Text-independent and Text-dependent according to nature of its work and categories of used features are described.

*Table 3: Differences Between Text-Independent And Text-Dependent In Writer Identification And Verification*

| | Text-Independent | Text-Dependent |
|---|---|---|
| **Nature of its work** | Text in the questioned document should not necessary be same as text in training process. | Text in the questioned document should be same as text in training process. |
| **Categories of used features** | Structural approach extremely involved to extract features from the characters shape. | Statistical and global writer identification techniques used widely to extract features from the text image. |

Journal of Theoretical and Applied Information Technology
30th November 2014. Vol. 69 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

### iv. DATASETS:

The availability of datasets is one of the most essential requirements in any research work. As such document analysis and recognition is not left out. A number of datasets comprising handwritten text samples have been developed over the years, mainly to support handwriting recognition, text segmentation and writer recognition tasks. Very recently, a number of standard datasets in different scripts and languages have been developed that allows researchers to evaluate their systems on the same databases for comparisons purposes. Some notable databases of Arabic handwriting samples along with their interesting statistics are summarized in Table 4.

*Table 4: Summary of notable Arabic handwritten text databases*

| DB Name | Language | Writers | Availability | Description |
|---|---|---|---|---|
| KHATT [18] | Arabic | 1000 | Upon request | 1000 forms, 2000 (random and fixed paragraphs) & free paragraphs |
| QUWI [19] | Arabic and English | 1017 | Upon request | |
| IFN/ENIT [20] | Arabic | 411 | Public | 26,459 images of Tunisian city names |
| PD100 [21] | Persian | 100 | Upon request | |
| BBN [22] | Arabic | 259 | Proprietary | |
| AHDB [23] | Arabic | 105 | Public | 10,000 words for check processing |
| Al Isra [24] | Arabic | 500 | Public | 37,000 words, 10,000 digits, 2500 signatures, 500 sentences |
| MADCAT [25] | Arabic | 400 | Public | 9693 handwritten pages from news groups, newswire and weblogs |
| CENPARMI [26] | Arabic | | Upon request | 3000 checks (Legal and courtesy amounts and digits) |
| AD/MADBase [27] | Arabic | 700 | Upon request | 700,000 digits |
| Alamri et al. [28] | Arabic | 328 | Upon request | 46,800 digits, 13,439 numerical strings, 21,426 letters, 11,375 words, 1640 special symbols |

The following section displays the most recent and important techniques that have been used to identify the authorship of handwritten samples in different databases. These are categorized into three divisions of local and global features that are locally and globally combined.

## 5. WRITER IDENTIFICATION: CURRENT STATE OF ART CLASSIFIED DEPENDING ON FEATURES TYPES

Writer recognition from Arabic handwritten documents have not been investigated as much as writer recognition from Latin or Chinese handwritten documents until the last few years. However, most of them share the same goal, which is to acquire individual features from a handwriting sample that can be used to identify authorship.

### i. Local Features

In [29], a novel method for writer identification has presented based on technique previously presented for character recognition, which is referred to as Oriented Basic Image Feature Columns (oBIF Columns). This can be used for writer identification. How this texture-based scheme can be enhanced by encoding a writer's style as the deviation from the mean encoding for a population of writers is a challenge. Delta encoding provides a more informative encoding than the texture-based encoding. The methods have been evaluated using the IAM dataset and by making entries into two top international competitions for assessing writer identification schemes. The authors demonstrate that the oBIF Column scheme on its own is sufficient to gain a performance level of 99% when tested using 300 writers from the IAM dataset, and in experiments conducted with the ICDAR 2012 Arabic datasets, the performance of the oBIF column system itself was comparatively high, achieved a score of 93.1% on the 204 writer dataset. However, the performance of the Δ encoding is considerably higher than with oBIF Columns.

In the work of [30] the presentation of Arabic writing can be seen as a form of new writing. This involves the presentation of Arabic writing in its

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

elementary constituents instead of it being alphabetic. The proposed technique divides the IFN/ENIT input document found in the database [20] into two parts: the first part contains letters and the second part is for the diacritics. The diacritics are extracted from the entire input image and used to calculate the LBP histogram for each diacritic before concatenating the obtained histograms used for handwriting feature studies. The results reveal that the proposed technique is highly dependable and handles Arabic handwriting better than previous methods. The success rate for identifying handwriting has increased to as high as 97.56%. The test has conducted with a total of 287 writers.

Studies by [31-33], proposed different systems to identify the writer of Arabic handwritten based on textural and allographic features. The function of the probability distribution was generated to extract the textural features and the closest neighbourhood classifier by applying the principle of measured distance for the allographic features by using the handwriting of 61 different writers as a sample to generate a codebook of 400 allographs in which the similar allograghs are put into separate feature. The best performance achieved was by the combination of some textural and allographic features which leads to a higher success rate of identifying the Top-1 of 88% and the Top-10 of 99% with an EER verification of about 5-6%. These approaches have been examined based on IFN/ENIT dataset [20] which include 350 writers containing 5 samples for each writer (where each sample is made up of 2 lines (about 9 words)).

Furthermore, Al-Maadeed et al. proposed two systems [34],[35] using an Arabic data set contain 32 000 images of Arabic text from 100 people with the system training set having 75% of the data and the remaining 25% has used for testing purposes.

[34] employed 16 words for identifying the authorship using Arabic text-dependent approach. The features extracted include the heights, lengths, and areas which were considered as edge-based directional features and it also included three edge-direction distributions include several different sizes and WED used as a classifier. The best result of 90% was obtained when 3 words were implemented in the top-10. The major drawback of this technique is its dependence on text and requires the use of only a small dataset for experimentation.

[35] used the same system as stated in [34] except employed the K-nearest neighbour classifier. The achievement of the new edge-based directional probability distributions and many features of Arabic writer identification were examined such as area, length, height, length from baseline to upper edge, and length from baseline to the lower edge. The success rate recorded for the top-10 writers was higher than 90%.

### ii. Global Features

Helli et **al**. proposed four systems [36-39] using PD100 as a testing dataset using 500 samples obtained from 100 writers in a single system and [36] used 70 persons in a written sample of 350.

In [36], an extended Gabor filter was employed in extracting the amount of percentage and styles of the curve of a writer's writing style and to extract the amount of percentage and style of the writers' word lines. Then the Weighted Euclidian Distance (WED) was employed to select the first three distances within the same class. The average rate of success to identify text-dependency was found to be 100% while the text-independency rate was 82%. However, the feature extraction using the XGabor filter could be useful to model the written documents characteristics however its accuracy measure is weak due to the challenges encountered in the classification and representation of the data. This has prompted the writers' of the current paper to combine the XGabor filter and the Gabor filter using several schemes for representing, classifying and identifying the data.

Furthermore [37], proposed new technique for classification which examines the order of the similarity sequence of the organized features using the Longest Common Subsequence (LCS) technique in comparing the similarity sequences. Gabor and XGabor filters were utilized in the extraction of the vector features. The average rate of success of identifying the writer by this method was 89%.

After that [38] has used the same systems as in [37] except the authors extracted the feature vectors using only the XGabor filters. The average success rate of identifying correctly the author was put at 95%.

[39] relied on Gabor and XGabor in extracting phrase features to elicit two sets of features from the dataset and created a Feature Relation Graph (FRG) employing the relationship existing within the extracted features. This graph was also referred to as DAFRG (Directed Acyclic FRG) and during the phase of classifying a dynamic algorithm it was utilized to find the optimum similarity index of any pairs in the DAFRG. The success rate for its ability to identify the correct author was 98%.

Journal of Theoretical and Applied Information Technology
30th November 2014. Vol. 69 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

### iii. Combining Global and Local Features

In Gazzah et al. three systems [40 - 42] relying on similar datasets consisting of 180 samples obtained from 60 different writers for testing purposes [40] by combining Global (2-D discrete wavelet transforms for analyzing the image text texture) and Local features (average line height, space between sub-words, inclination of the ascenders and features extracted from dots) were used to extract the features were extracted. Modular Multilayer Perceptron (MMP) was used. On average the success rate for identifying the author was 95.68%.

[41] is similar to [40] by the addition of an extra global feature which is called Entropy. The success rate on average was 94.73%. [42] is quite similar to [41] by the addition of an extra classifier called SVM. The success rate recorded on the average was 93.76%.

Siddiqi and Vincent [43] employed the same idea of frequent writing patterns to characterize a writer but instead of using graphemes, a scale of observation was employed. The authors first binarized the handwriting and divided the text into small fragments by positioning small square windows on the text. Similar fragments were then grouped together into clusters (codebook) [44]. The authors first used writer specific codebooks, to generate a unique but different codebook for each writer and later shifted to a general codebook where a single codebook was generated using a set of writing samples thus making the approach similar to [45]. To complete the codebook based features, the authors also introduced a set of contour based features to capture the curves and the orientation information contained in the write-up [46]. The codebook and contour features were combined together [47], and evaluated on different datasets such as IAM (650 writers), RIMES (375 writers) and the Arabic handwriting IFN/ENIT database [20] and all the parameters were able to obtain some very promising outcomes (for 100 sampled writers the rate of identification was 92% with an EER of 2.94%).

Furthermore, Djeddi et al. proposed two systems [48], [49] for text-independent writer recognition from Arabic handwritten documents while [48] proposed a novel approach for text-independent writer recognition from Arabic handwritten documents. To characterize the styles of the handwriting of several different writers involved in the evaluation of our approach, the authors have used two texture methods based on edge hinge features and run-lengths features. The advantage of this method is shown by experimenting on the classification of (1750) images of Arabic handwriting from (275) unique writers in IFN/ENIT database [20]. The identification rate is increased by combining the run-lengths features with edge hinge feature to achieve an identification rate of 93.53% in Top 1, 98.47% in Top 5 and 99.13% in Top 10 and an EER of 4.78%.

[49] analyzed the sensitivity of codebook-based writer identification systems of the patterns in the codebook. A number of data sets in a variety of languages were used for this purpose. [49] first showed that a codebook generated from a script different than those of writings under study achieved identification rates substantially approaching those of the classical codebook based methods. A number of data sets with handwritten samples in Arabic, French, English, German, Urdu and Greek are considered to evaluate this technique. The best writer identification rates for MSHD-Arabic (87 writers) on codebook based on handwritten music scores samples from the CVC data set [50] is 91.19%.

To the best of our knowledge, the most important methods in Arabic writer identification are reported in this study. Table 5 summarizes the Writer Identification Methods on Arabic and Persian Languages.

Journal of Theoretical and Applied Information Technology
30th November 2014. Vol. 69 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Table 5: Writer Identification Methods*

| System | Sample Space | Features | Classification Methodology | Accuracy | Language | Text-Dependent / Independent |
|---|---|---|---|---|---|---|
| Al-Ma'adeed et al. [34],[35] | 100 writers(32000 Arabic text images (16 different types)) | Height area, length and Edge –direction distribution | WED classifier K-nearest neighbor -trained with 75% -tested with 25% | Top-10: 90% | Arabic | text-dependent |
| Shahabi and Rahmati,[51] | 25 persons – Arabic/ Persian handwriting | Texture image Gabor function | Euclidean Distance | 92% | Arabic/ Persian | text-independent |
| Newell, et al. [29] | 204 writer (ICDAR 2012) | Delta encoding oriented Basic Image Features (oBIF Columns) | Nearest neighbor Kernel Principal Component Analysis and Support Vector Machines (SVMs) | 95.3% | Arabic | text-independent |
| Siddiqi and Vincent [47] | 100 writers IFN/ENIT [20] | contour-based orientation and curvature features | nearest neighbor | 92% | Arabic | text-independent |
| Djeddi, et al. [48] | 275 writers 1750 handwriting IFN/ENIT [20] | edge hinge features and run-lengths features | nearest-neighbor classification | Top 10: 99.13% | Arabic | text-independent |
| Djeddi, et al. [49] | 87 writers MSHD- Arabic | global and local features | K Nearest Neighbor classifier | 91.19% | Arabic | text-independent |
| Bulacu et al. [33] | 350 writers with 5 samples per writer IFN/ENIT [20] | combining some textural and allographic features | nearest neighborhood classifier | Top-1 88% and Top 10: 99% | Arabic | text-independent |
| Lutf, M et al. [30] | 287 writers IFN/ENIT [20] | local binary pattern histogram (LBP) for each diacritic | K-nearest Neighbor | 97.56% | Arabic | Not Mentioned |
| B. Helli et al. [39] | 100 writers (PD100 dataset), 500 samples [21] | Gabor & XGabor filters | Employed Dynamic Algorithm for obtaining the optimum similarity | 98% | Persian | text-independent |
| B. Helli et al. [37] | 100 writers (PD100 dataset), 500 samples [21] | Same features in [39] | proposed a novel technique for classification that measures the similarity sequence by depending on the Longest Common Subsequence (LCS) | 89% | Persian | text-independent |
| B. Helli et al. [38] | PD100 dataset which include 500 samples from 100 writers [21] | Point-based (speed, acceleration, vicinity linearity, vicinity slope), stroke-based (duration, time to next stroke, number of points, number of up strokes, etc.). | They presented a classifier that is based on LCS (longest common subsequence) | 95% | Persian | text-independent |
| S. Gazzah et al. [40] | 180 samples from 60 writers | combination 2-D Discreet wavelet transforms-lifting & Statistical features | Modular Multilayer Perceptron | 95.68% | Arabic | Not Mentioned |
| S. Gazzah et al. [41] | 180 samples from 60 writers | Same features in [40] by the addition of an extra global feature called Entropy. | Modular Multilayer Perceptron | 94.73% | Arabic | Not Mentioned |

Journal of Theoretical and Applied Information Technology
30th November 2014. Vol. 69 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

## 6. DISCUSSION AND CONCLUSION

The survey of the recent methods for off-line text-independent and text-dependent Arabic writer identification is presented in this paper. This paper provides a summary of existing methods and their main challenges in the area. Finding the best and most appropriate writing feature sets to represent handwriting image and the best practice off-line text-independent and text-dependent Arabic writer identification are challenging issues. In Arabic language the writer specific texture features using multichannel Gabor filtering, edge distribution and Gray-Scale Co-occurrence Matrices are common. Also the studies are carried out in other languages such as Persian. Combinations of some textural and allographic features as well as statistical measures, multiple-channel (Gabor) filters, XGabor are carried out to obtain the individuality of the writers and better accuracy. Also studies are based on the features used in other languages. Current methods based on a universal codebook are generally popular and efficient in terms of computational cost, however, a new codebook is to be generated if the script changes. From this we understand that features must be selected based on the characteristic features of each language. Future researchers need to finally develop a recognition system that is practical for widespread use. There is a long way to go in order to fully develop a system for writer recognition. There is a lot of scope for further research in this area.

## REFERENCES

[1]. Bunke, H. (2003). Recognition of cursive Roman handwriting: past, present and future. Proceedings of the 2003 Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on. 3-6 Aug. 2003. 448-459.

[2]. Guyon, I., Schomaker, L., Plamondon, R., Liberman, M. and Janet, S. (1994). Unipen project of on-line data exchange and recognizer benchmarks. In ICPR'94: In Proceedings of the 12th International Conference on Pattern Recognition, 29–33.

[3]. Marius B. and Lambert S. (2007). Text-Independent Writer Identification and Verification Using Textural and Allographic Features. , IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, No. 4.

[4]. Palhang, M. and Sowmya, A. (1999). “Feature Extraction: Issues, New Features, and Symbolic Representation”, Third International Conference, VISUAL'99, Amsterdam, Vol. 1614. pp. 418 – 427.

[5]. Gunter, S.and Bunke, H. (2002). “Creation of Classifier Esembler for Handwritten Word Recognition Using Feature Selection Algorithms”, IEEE International Workshop on Frontiers in Handwriting Recognition. 2002.

[6]. Jain, A., Duin, R. and Mao, J. C. (2000), “Statistical Pattern Recognition : A Review”, IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 22. No.1. pp. 4-37.

[7]. Jain, A. and Zongker D. (1997), “Feature Selection : Evaluation, Application and Small Sample Performance”, IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 19. No. 2. pp. 153 – 158.

[8]. Zhang, P., Chen, L. and Alex, C.-K. (2000), “A Novel Hybrid Classifier for Recognition of Handwritten Numerals”, IEEE, 0-7803-6583-6, pp. 2709-2714.

[9]. Khotanzad, A. and Lu, J.-H. (1990). “Classification of Invariant Image Representations Using a Neural Network”. IEEE Transaction on Acoustics, Speech and Signal Processing, Vol. 38. No. 6. pp. 1028-1038.

[10]. Zhenyu, H., Yuan, Y. - T. and Xinge, Y. (2005). “A Contourlet-based Method for Writer Identification”, IEEE International Conference on Systems, Man and Cybernetics, Vol. 1. pp. 364 – 368.

[11]. He, Z.-Y., and Tang, Y.-Y. (2004). “Chinese Handwriting-based Writer Identification by Texture Analysis”. Proceedings of International Conference of Machine Learning and Cybernetics, Vol. 6. pp. 3488 – 3491.

[12]. Zois, E.N. and Anastassopoulos, V. (1996). “Methods for Writer Identification”, Proceedings of the Third IEEE International Conference on Electronics, Circuits, and Systems (ICECS '96), Vol. 2. pp. 740 – 743.

[13]. Shen, C., Ruan, X.-G. and Mao, T.-L. (2002). “Writer Identification Using Gabor Wavelet”, Proceedings of the 4th World Congress on Intelligent Control and Automation, Vol. 3. pp. 2061 – 2064.

[14]. Kun, Y., Yunhong. W. and Tieniu, T. (2004). “Writer Identification Using Dynamic Features”, Biometric Authentication: First International Conference, ICBA 2004, Hong Kong, China, pp. 512 – 518.

[15]. He, Z.-Y., Chen, Q.-H. and Chen, D.-F. (2002). “A Neural Network Expert System for

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Chinese Handwriting-based Writer Identification", Proceedings of International Conference of Machine Learning and Cybernetics, Vol. 4. pp. 2194 – 2196.

[16]. Said, H.E.S., Tan, T.N. and Baker, K.D. (2000). "Writer Identification Based on Handwriting", Pattern Recognition, Vol. 33. Pp.149-160.

[17]. Schlapbach, A. and Bunke, H. (2004a). "Using HMM Based Recognizers for Writer Identification and Verification", In Proc. 9th Int. Workshop on Frontiers in Handwriting Recognition, pp. 167–172.

[18]. Mahmoud, S. A., Ahmad, I., Al-Khatib, W. G., Alshayeb, M., Tanvir Parvez, M., Märgner, V., & Fink, G. A. (2014). KHATT: An open Arabic offline handwritten text database. Pattern Recognition, 47(3), 1096-1112.

[19]. Al-Máadeed, S., Ayouby, W., Hassaïne, A., & Al'Jaam, J. M. (2012, September). QUWI: An Arabic and English Handwriting Dataset for Offline Writer Identification. In ICFHR (pp. 746-751).

[20]. M. Pechwitz, S. S. Maddouri, V. Märgner, N. Ellouze, H. Amiri, IFN/ENIT database of handwritten Arabic words, in: Seventh Colloque International Francophone sur l'Ecrit et le Document, CIFED 2002, 2002, pp. 129–136.

[21]. Helli, B. and Moghaddam, M. E. (2008). A Text-Independent Persian Writer Identification System Using LCS Based Classifier. Proceedings of the 2008 Signal Processing and Information Technology, 2008. ISSPIT 2008, IEEE International Symposium, 203-206.

[22]. H. Cao, R. Prasad, and P. Natarajan. Improvements in hmm adaptation for handwriting recognition using writer identification and duration adaptation. In Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on, pages 154 –159, nov. 2010.

[23]. S. Al-Maadeed, D. Elliman, and C. Higgins. A Dataset for Arabic Handwritten Text Recognition Research. The International Arab Journal of Information Technology, 2004.

[24]. N. Kharma, M. Ahmed, R. Ward, A new comprehensive database of handwritten Arabic words, numbers, and signatures used for OCR testing, in: Canadian Conference on Electrical And Computer Engineering, 1999, pp. 766–768.

[25]. R. Al-Hajj, C. Mokbel, L. Likforman-Sulem, Combination of HMM-based classifiers for the recognition of Arabic handwritten words, in Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), 2007,vol. 2, pp. 959–963.

[26]. Y. Al-Ohali, M. Cheriet, C.Y. Suen, M. B, Databases for recognition of handwritten Arabic cheques, Pattern Recognition 36 (1) (2003) 111–121.

[27]. E. A. El-Sherif, S. Abdelazeem, A two-stage system for Arabic handwritten digit recognition tested on a new large database, in: International Conference on Artificial Intelligence and Pattern Recognition, 2007, pp. 237–242.

[28]. H. Alamri, J. Sadri, C. Y. Suen, N. Nobile, A novel comprehensive database for arabic off-line handwriting recognition Huda Alamri," in: Eleventh International Conference on Frontiers in Handwriting Recognition, 2008.

[29]. Newell, Andrew J., and Lewis D. Griffin. "Writer identification using oriented Basic Image Features and the Delta encoding." Pattern Recognition 47.6 (2014): 2255-2265.

[30]. Lutf, M., You, X., & Li, H. (2010, August). Offline arabic handwriting identification using language diacritics. In Pattern Recognition (ICPR), 2010 20th International Conference on (pp. 1912-1915). IEEE.

[31]. Bulacu, M., Schomaker, L., Vuurpijl, L. (2003) Writer identification using edge-based directional features, in: Seventh International Conference on Document Analysis and Recognition (ICDAR).

[32]. Bulacu, M., Schomaker, L.(2007) Text-independent writer identification and verification using textural and allographic features, IEEE Transactions on Pattern Analysis and Machine Intelligence(PAMI)29(4) 701–717 Special Issue—Biometrics: Progress and Directions.

[33]. Bulacu, M., Schomaker, L., Brink, A. (2007) Text-independent writer identification and verification on offline Arabic handwriting, in: Ninth Conference on Document Analysis and Recognition (ICDAR).

[34]. Al-Ma'adeed, S., Mohammed, E., AlKassis, D., Al-Muslih, F. (2008) Writer identification using edge-based directional probability

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

distribution features for Arabic words, in: IEEE/ACS International Conference on Computer Systems and Applications (AICCSA), 582—590.

[35]. Al-Maadeed, S. (2012). Text-dependent writer identification for Arabic handwriting. Journal of Electrical and Computer Engineering, 2012, 13.

[36]. B. Helli and M. Moghadam, "Persian Writer Identification Using Extended Gabor Filter," in Image Analysis and Recognition. vol. 5112, A. Campilho and M. Kamel, Eds., ed: Springer Berlin / Heidelberg, 2008, pp. 579-586.

[37]. B. Helli and M. E. Moghaddam, "A writer identification method based on XGabor and LCS," IEICE Electron. Express, vol. VOL .6, pp. 623-629, 2009.

[38]. B. Helli and M. E. Moghaddam, "A Text-Independent Persian Writer Identification System Using LCS Based Classifier," in Signal Processing and Information Technology, 2008. ISSPIT 2008. IEEE International Symposium on, 2008, pp. 203-206.

[39]. B. Helli and M. E. Moghaddam, "A text-independent Persian writer identification based on feature relation graph (FRG)," Pattern Recogn., vol. 43, pp. 2199-2209, 2010.

[40]. S. Gazzah and N. E. Ben Amara,"Arabic Handwriting Texture Analysis for Writer Identification Using the DWT-Lifting Scheme,"in Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on, 2007, pp. 1133-1137.

[41]. S. Gazzah and N. Essoukri Ben Amara, "Writer Identification Using Modular MLP Classifier and Genetic Algorithm for Optimal Features Selection,"in Advances in Neural Networks - ISNN 2006. vol. 3972, J. Wang, et al., Eds., ed: Springer Berlin / Heidelberg, 2006, pp. 271-276.

[42]. S. Gazzah and N. E. B. Amara, "Neural Networks and Support Vector Machines Classifiers for Writer Identification Using Arabic Script,"The International Arab Journal of Information Technology, vol. VOL. 5, pp. 92-101 2006.

[43]. Siddiqi, I. A. and Vincent, N. (2007). Writer identification in handwritten documents. Proceedings of the 2007 Curitiba, 108-112.

[44]. Siddiqi, I. A. and Vincent, N. (2008b). combining global AND local features for writer identification. Proceedings of the 2008b The 11th International Conference on Frontiers in Handwriting Recognition Concordia University, Montréal, Québec, Canada,

[45]. Bulacu, M. and Schomaker, L. (2007b). Text-independent writer identification and verification using textural and allographic features. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(4), 701-717.

[46]. Siddiqi, I., Khurshid, K. and Vincent, N. (2011). Feature relevance analysis for writer identification. Proceedings of the 2011 San Francisco, CA.

[47]. Siddiqi, I. and Vincent, N. (2010). Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. Pattern Recognition, 43(11), 3853-3865.

[48]. Djeddi, C., Souici-Meslati, L., & Ennaji, A. (2012). Writer recognition on arabic handwritten documents. In Image and Signal Processing (pp. 493-501). Springer Berlin Heidelberg.

[49]. Djeddi, C., Siddiqi, I., Souici-Meslati, L., & Ennaji, A. (2013, August). Codebook for Writer Characterization: A Vocabulary of Patterns or a Mere Representation Space? In Document Analysis and Recognition (ICDAR), 2013 12th International Conference on (pp. 423-427). IEEE.

[50]. Fornés, A. Dutta, A. Gordo, J. Lladós, "CVC-MUSCIMA: a ground truth of handwritten music score images for writer identification and staff removal", In International Journal on Document Analysis and Recognition, September 2012, Volume 15, Issue 3, pp 243-251.

[51]. Shahabi, F., Rahmati, M. (2006) Comparison of Gabor based features for writer identification of Farsi/Arabic handwriting, in: 10th International Workshop on Frontiers in Handwritten Recognition (IWFHR).